

ABC HealthCare Case Study
Internal Audit - Payment Process at ABC Healthcare
Leveraging Data Science and AI to strengthen Payment Oversight



Background:

ABC HealthCare are a global pharmaceutical business specialising in cancer research. It has 206 independent research teams (RT) operating with universities and research organisations throughout the world. Each team are responsible for their own areas of research and budget, however services such as legal and finance are provided and managed centrally by ABC HealthCare.

ABC Healthcare's payments team are based in Thailand and receive over 200,000 payment requests a year from research teams around the world. They are responsible for checking requests and processing payments to suppliers on behalf of the research teams.

As part of the organisations ongoing commitment to integrity and transparency within their financial operations, the Payment Manager is tasked with overseeing and authorising payments in collaboration with Payment Director.

Submission and payment authorisation follow a delegated authority's structure as explained below:

Invoice Amount	Submission authoriser	Payment Authoriser
< \$ 1000	Submitter	Payment Analyst
< \$ 5000	Submitter's Manager	Payment Manager
> \$ 5000	Submitter's Director	Payment Director

Recently, a whistleblower report has brought to light concerns regarding potentially fraudulent activities within our payments process. The report outlined several examples of irregularities, including:

- Research teams submitting fraudulent invoices,
- Invoices not matching claims and
- Instances of employee fraud within the finance team.

A preliminary investigation conducted by the leadership team indicates that there is sufficient cause to warrant a more comprehensive audit. Specifically, it has been recommended that we employ advanced data analytics techniques to thoroughly examine both the payments data and the underlying processes.

As a member of the Data & AI team embedded within the internal audit function, you are at the forefront of this analytics-driven review. Your responsibility is to apply **Artificial Intelligence (AI) and Data Science** skills to support the audit and provide actionable insights to your fellow internal auditors throughout this critical project. Your work will be instrumental in identifying potential control weaknesses, detecting anomalies or suspicious patterns, and ultimately strengthening the integrity of organisations payments process.

Extracting the supporting claims documents is difficult due to old and legacy based finance systems being used by the business and the location of the business function, however the business team has been able to provide some sample records that can support the audit objective.

A brief overview of the sample records/files along with data mapping has been provided below that can help support your analytics work:

Dataset Overview

For the case study you will be using 5 datasets as listed below:

1. Payments Master: Contains 2000 sample records of Payments processed
2. Research Team Master: Contains the information of the Research teams
3. Research Team members Master: Contains information on members of the Research teams
4. Fraud Cases Master (Training Dataset): Like payments data file with an additional fraud flag

You have been provided with one .zip files as follows:

- Payments Information.zip: Containing all relevant information with datasets 1- 4.

1. Payments Master (.csv):

- **Date received:** When the invoice was received.
- **Date of invoice:** When the invoice was issued.
- **Date of authorisation:** When the invoice was approved.
- **Payment due date:** When the payment is due.
- **Date of payment:** When the payment was made.
- **Research team:** Team responsible for the expense.
- **Submitted by:** Person who submitted the invoice.
- **Authorised by:** Person who approved the invoice.
- **Payment authoriser:** Person who approved the payment.
- **Invoice number:** Unique identifier for the invoice.
- **Description of spend:** What the invoice is for.
- **Invoice value:** Amount on the invoice.
- **Payment amount:** Amount paid.
- **Payment Status:** e.g., Paid, Pending.
- **Type of expense:** e.g., Lab suppliers, Consulting, etc.
- **Company:** Vendor or service provider.
- **phone_number:** Contact number for the company or submitter.
- **email:** Contact email.

2. Teams' Master (.csv):

- **Research team:** Name or identifier of the research group.
- **Director:** Person leading the team.
- **Location:** Geographic or institutional location.
- **Affiliation:** Associated institution or department.
- **Research type:** Field of research (e.g., Basic Research', 'Translational', etc.)
- **Annual budget:** Total yearly funding allocated to the team.
- **Item type:** Specific budget item (e.g., Lab suppliers, Consulting, etc.).
- **Item budget:** budget associated with the item.
- **Comments:** Additional notes or context.

3. Team Members Master (.csv):

- **Team:** The name or identifier of the team.
- **Location:** Where the team or individual is based.
- **Name:** The name of the team member.
- **Role:** The position or function of the person within the team.

4. Fraud Cases Master (Training Dataset):

A .csv file containing same data fields template as **1. Payment Master**, with different data values, and an additional column stated as Fraud flag which Indicates:

- **1** if it was identified as a fraud case, and
- **0** if not.

Instructions and Case Study Tasks

This case study is structured in 2 sections, both of which are mandatory to complete

1. Section 1: Descriptive and Diagnostic Analytics of datasets
2. Section 2: Use of Data Science/ AI or Machine Learning Techniques. You may focus upon **supervised** or **unsupervised** learning if you prefer a particular approach. However, you may be asked about both approaches at interview.

Submission Requirements: It is expected that you work on the case study, generating a brief presentation with your audience being the Group Internal Audit Leadership Team (10 slides maximum including both sections) to present the most important outcomes and the style of presentation should be at an Executive level. This presentation should be sent to your recruiting coordinator at least **two working days** before your technical interview is scheduled.

Also, as part of your submission, feel free to use the tool of your choice for performing the analysis and completion of the case study but please consider sharing any files such as scripts, notebooks, approach, apps, etc. which you think can support or explain your analytics work.

Finally, you can document or explain the process and any insights you feel the audit team would find useful, how you performed your analysis, any assumptions, and conclusions. For example: Are there any missing data fields which may have helped you identify fraud?

Section 1: Descriptive and Diagnostic Analytics

A. Using PowerBI or an appropriate visualisation package:

- Any analysis and / or visualisations which may suggest or help identify fraud such as - Identify the research teams with the highest spend; Track volume and time trends by category and location; average payment time by expense type, value and location; comparing each team spend against their budget.

B. What impacts the time taken to pay the invoice/ request:

Hypothesis: All invoices take approximately the same amount of time, plus or minus one day.

- Task: Perform regression or similar modelling analysis to predict how long it will take to make a payment.

C. Identify statistical outliers which may require investigation by auditors/management:

Hypothesis: All invoices are random and there is no pattern which allows the team to predict and flag potential fraud.

- Task: Identify statistical outliers to select samples for the audit team to further investigate that indicate potential fraud cases. For example – Using Z scores or similar techniques.

Section 2: Machine Learning and AI

Submission Requirements: For Section 2 – you will be required to complete both Part A (Machine Learning) and Part B (Use of AI) tasks of the case study.

You will be required to demonstrate and explain the approach adopted to build the predictive model along with providing any appropriate evaluation techniques use to determine the accuracy of the model such as precision, confusion matrix or perplexity.

Your presentation should contain key elements of code, outputs, evaluation to support any conclusions and recommendations. Your conclusions and recommendations need to be presented at an Executive level. Your readers are likely to be a mix of analytics technical and non-technical individuals.

A. Machine Learning: Using an appropriate R/Python package(s):

1. Build and explain a supervised model that predicts which payment requests could be fraudulent.

And/or

2. Build and explain an unsupervised model that predicts which payment requests could be fraudulent.

And

B. Use of AI:

Demonstrate the approach of how you would use an AI tool of your choice that supports an AI generated Audit report from your analysis and any key Insights/findings from work performed across Sections 1 and 2.

Whilst the creation of a complete audit report is not expected during the interview round, you should be able to demonstrate your knowledge and experience in performing such a task.

Finally, summarise your approach alongside any assumptions and limitations in your submissions.

Your conclusions and recommendations need to be presented at an Executive level. Your readers are likely to be a mix of analytics technical and non-technical individuals.